

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2001, 10:52:55 ; Search time 30.06 Seconds

(without alignments)
31.624 Million cell updates/sec

Title: US-09-288-719-2

Perfect score: 75

Sequence: 1 GGGSGGGRASGGGS 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 segs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database :

1: PIR66:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	71	94.7	233	2 JC5322	p53 specific singl
2	65	86.7	643	1 KRNU2	keratin I, type II
3	62	82.7	206	2 I53066	gene M-twist prote
4	61	81.3	102	2 D71415	hypothetical prote
5	60	80.0	401	1 OZ2QAC	clitumsporozoite p
6	60	80.0	465	2 G02738	FRAC-4 - human
7	60	80.0	592	2 E82759	endo-1,4-beta-gluc
8	60	80.0	779	2 A35006	cell surface glyco
9	60	80.0	1433	2 A46053	bullous pemphigoid
10	59	78.7	1585	2 T31611	hypothetical prote
11	59	78.7	136	2 T29282	hypothetical prote
12	59	78.7	221	2 T04592	hypothetical prote
13	59	78.7	291	1 S31415	glycine-rich cell
14	59	78.7	316	2 T20497	hypothetical prote
15	59	78.7	558	2 A33616	heterogeneous ribo
16	59	78.7	561	2 A31994	keratin 10, type I
17	58	77.3	214	1 KRNU2	keratin 10, type I
18	58	77.3	268	2 A56446	glycine-rich prote
19	58	77.3	284	2 S74256	ig heavy chain V r
20	58	77.3	330	2 S74255	homeotic protein s
21	58	77.3	333	2 A39065	homeotic protein E
22	58	77.3	427	2 A32372	female-specific do
23	58	77.3	435	2 T15143	hypothetical prote
24	58	77.3	481	2 A35628	loricrin - mouse
25	58	77.3	495	2 B71360	hypothetical prote
26	58	77.3	549	2 B32372	male-specific doub
27	58	77.3	877	2 T43449	hypothetical prote
28	58	77.3	877	2 T43449	hypothetical prote
29	57	76.0	239	2 S49193	CCR 101 protein -

30	57	76.0	265	2 A40649	hypothetical prote
31	57	76.0	266	1 CIRBL	calpain (EC 3.4.22
32	57	76.0	268	2 S09860	hypothetical prote
33	57	76.0	271	2 S34666	glycine-rich prote
34	57	76.0	312	2 T25048	hypothetical prote
35	57	76.0	396	2 T49109	glycine-rich prote
36	57	76.0	431	1 WJHU2G	homeotic protein H
37	57	76.0	433	2 S20963	homeotic protein H
38	57	76.0	500	2 T22068	hypothetical prote
39	57	76.0	575	1 S11485	hypothetical prote
40	57	76.0	586	2 T26667	RNA helicase - fru
41	57	76.0	1454	2 T13709	hypothetical prote
42	56	74.7	80	2 T10550	diacylglycerol kin
43	56	74.7	157	1 S14857	hypothetical prote
44	56	74.7	157	2 S04536	glycine-rich prote
45	56	74.7	165	1 KNR2G1	embryonic abundant
					glycine-rich cell

ALIGNMENTS

RESULT 1
JC5322
p53 specific single-chain antibody Pab421 - human
C:Species: Homo sapiens (man)
C:Date: 15-May-1997 #sequence_revision 15-May-1997 #text_change 18-Jul-1997
C:Accession: JC5322
R:Jannot, C.B.; Hynes, N.E.
Biochem. Biophys. Res. Commun. 230, 242-246, 1997
A:Title: Characterization of scFv-421, a single-chain antibody targeted to p53.
A:Reference number: JC5322; MUID:97168950
A:Accession: JC5322
A:Molecule type: mRNA
A:Residues: 1-233 <JAN>
A:Experimental source: hydrioloma cell
C:Comment: This protein specifically binds the tumor suppressor protein p53. It resto

Query Match 94.7%; Score 71; DB 2; Length 233;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGSGGGRASGGC 13

DB 110 GGGSGGGRASGGC 122

RESULT 2

KRNU2
keratin I, type II, cytoskeletal - human
N:Alternate names: 67K type II epidermal keratin; cytokeratin 1

C:Species: Homo sapiens (man)

C:Date: 04-Dec-1986 #sequence_revision 22-Oct-1999 #text_change 10-Dec-1999

C:Accession: A22940; A02950; A43342

R:Johnson, L.D.; Idler, W.W.; Zhou, X.M.; Roop, D.R.; Steinert, P.M.

Proc. Natl. Acad. Sci. U.S.A. 82, 1896-1900, 1985

A:Reference number: A22940; MUID:85166239

A:Accession: A22940

A:Molecule type: DNA

A:Residues: 1-643 <JOH>

A:Cross-references: GB:M98776; GB:M11215; GB:M11845; GB:M11846; NID:g1843461

A:Note: translation of Initiator Met is not shown

R:Steinert, P.M.; Parry, D.A.D.; Idler, W.W.; Johnson, L.D.; Steven, A.C.; Roop, D.R.

J. Biol. Chem. 260, 7142-7149, 1985

A:Title: Amino acid sequences of mouse and human epidermal type II keratins of M-r 67

lact filament subunits.

A:Reference number: A92535; MUID:85207740

A:Accession: A02950

A:Molecule type: mRNA

A:Residues: 151-183, 'K', 185-199, 'M', 201-204, 'K', 206-236, 'S', 238-239, 'R', 241-356, 'Y', 3

'S', 638-643 <STE>

A:Cross-references: GB:M10938; NID:g186787; PIDN:AAA36153.1; PID:g386854

A:Experimental source: tissue neonatal foreskin

A:Note: the authors translated the codon CUG for residue 476 as Met
R:Chipev, C.C.; Korge, B.P.; Markova, N.; Bale, S.J.; Digiovanna, J.J.; Compton, J.G.;
Cell 70, 821-828, 1992
A:Title: A leucine----proline mutation in the H1 subdomain of keratin 1 causes epidermol
A:Reference number: A43342; MUID:92386601
A:Accession: A43342
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 144-146, 'P', 148-159, 'P', 161-183, 'K', 185-186 <CHI>
A:Cross-references: GB:M68776; GB:M11215; GB:M11845; GB:M1846; NID:g1843461
A:Note: sequence extracted from NCBI backbone (NCBI:112784)
C:Comment: The cytoskeletal and microfilillar keratins are classified into two types, t
atin IF protein subunit appears to be a heterotrimer of two type I and two type II pro
C:Comment: Keratin 1 is expressed in terminally differentiating epidermis.
C:Genetics:
A:Gene: GDB:KRL1
A:Cross-references: GDB:128198; OMIM:139350
A:Map position: 12q11-12q13
A:Note: defects in this gene may result in epidermolytic hyperkeratosis
C:Complex: heterotrimer of two type I, usually keratin 10 (see PIR:KRH0), and two ty
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; heterotrimer; intermediate filament
F:4-179/Domain: head <HED>
F:4-143/Region: E1 and V1 subdomains
F:14-179/Region: H1 subdomain
F:180-492/Domain: rod <ROD>
F:180-214/Region: coil 1A
F:215-226/Region: linker 1
F:227-327/Region: coil 1B
F:328-344/Region: linker 12
F:345-363/Region: coil 2A
F:364-371/Region: linker 2
F:372-492/Region: coil 2B
F:430/Region: stutter
F:493-643/Domain: tail <END>
F:493-512/Region: H2 subdomain
F:513-643/Region: V2 and E2 subdomains

Query Match 86.7%; Score 65; DB 1; Length 643;
Best Local Similarity 85.7%; Pred. No. 1.3;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGSGGGRASGGS 14
DB 594 GGGSGGGRASGGS 607

RESULT 3
153066
gene M-twist protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C:Accession: I53066; I66795
R:Wolf, C.; Thisse, C.; Stoetzel, C.; Thisse, B.; Gerlinger, P.; Perrin-Schmitt, F.
Dev. Biol. 143, 365-373, 1991
A:Title: The M-twist gene of Mus is expressed in subsets of mesodermal cells and is clos
A:Reference number: I53066; MUID:91122450
A:Accession: I53066
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-206 <RES>
A:Cross-references: GB:M63649; NID:g202243; PIDN:AAA0514.1; PID:g202244
A:Accession: I66795
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-35, 'R', 37-90, 'P', 92-206 <RE2>
A:Cross-references: GB:M63650; NID:g202245; PIDN:AAA0515.1; PID:g202246
C:Genetics:
A:Gene: M-twist

Query Match 82.7%; Score 62; DB 2; Length 206;

Best Local Similarity 78.6%; Pred. No. 1;
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGSGGGRASGGS 14
DB 90 GGGSGGGRASGGS 103

RESULT 4
D71415
hypothetical protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
C:Accession: D71415
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
P.; Medler, H.; Medler, E.; Manbut, R.; Meltzenger, T.; Pohl, T.M.; Terry, N.; G
avanagh, T.; Hempel, S.; Kotter, P.; Ertlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Pulgömen
ehoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chalvatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A:Reference number: A71400; MUID:98121113
A:Accession: D71415
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-102 <BEV>
A:Cross-references: GB:Z97338; NID:g2244870; PID:e326910; PID:g2244873
C:Genetics:
A:Map position: 4COP9-4G3845

Query Match 81.3%; Score 61; DB 2; Length 102;
Best Local Similarity 78.6%; Pred. No. 0.73;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGSGGGRASGGS 14
DB 50 GGGSGGGRASGGS 63

RESULT 5
OZZOAC
circumsporozoite protein precursor - Plasmodium cynomolgi (strain Gombak)
N:Alternate names: major sporozoite surface antigen
C:Species: Plasmodium cynomolgi
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 07-Nov-1997
C:Accession: E26255
R:Galinski, M.R.; Arnot, D.E.; Cochrane, A.H.; Barnwell, J.W.; Nussenzweig, R.S.; Ene
Cell 48, 311-319, 1987
A:Title: The circumsporozoite gene of the Plasmodium cynomolgi complex.
A:Reference number: A90889; MUID:87102878
A:Accession: E26255
A:Molecule type: DNA
A:Residues: 1-401 <CAL>
C:Comment: There are three distinct regions in the mature circumsporozoite protein, t
obic membrane anchoring sequence.
C:Comment: There are 17 tandem copies of the 11-residue repeat D/G-G-A-A-A-G-G-G-G-
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
C:Keywords: sporozoite; surface antigen; tandem repeat
F:1-19/Domain: signal sequence #status predicted <Sig>
F:20-401/Product: circumsporozoite protein #status predicted <MNT>
F:88-278/Region: 11-residue repeats
F:326-379/Domain: thrombospondin type 1 repeat homology <THRL>

Query Match 80.0%; Score 60; DB 1; Length 401;
Best Local Similarity 71.4%; Pred. No. 3.1;
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGSGGGRASGGS 14
|||||:|:|:|:

R.Steward, C.
 Submitted to the EMBL Data Library, September 1999
 A:Reference number: 221047
 A:Accession: T31611
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1585 <WIL>
 A:Cross-references: EMBL:AL117200; NID:e1549770; PIDN:CAB55050.1; CESP:Y50E8A.g
 C:Genetics:
 A:Experimental source: clone Y50E8A
 A:Gene: CESP:Y50E8A.g
 A:Introns: 25/3; 60/1; 133/2; 217/3; 270/3; 337/2; 400/1; 746/2

Query Match
 Best Local Similarity 84.6%; Pred. No. 10;
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGGGSGGRASGGG 13
 DB 452 GGGGSGGRASGGG 464

RESULT 11
 T29282
 hypothetical protein C34D4.11 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T29282
 R:Du, Z.; Le, T.T.
 Submitted to the EMBL Data Library, May 1996
 A:Description: The sequence of C. elegans cosmid C34D4.
 A:Reference number: 220600
 A:Accession: T29282
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-136 <DU>
 A:Cross-references: EMBL:U58755; PIDN:AAB00696.1; GSPDB:GN00022; CESP:C34D4.11
 C:Genetics:
 A:Gene: CESP:C34D4.11
 A:Map position: 4
 A:Introns: 20/1; 66/1; 98/1; 116/1
 C:Superfamily: Arabidopsis glycine-rich protein 3

Query Match
 Best Local Similarity 78.7%; Score 59; DB 2; Length 136;
 Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGGGSGGRASGGG 13
 DB 97 GGGGSGGRASGGG 109

RESULT 12
 T04592
 glycine-rich cell wall structural protein homolog F23E13.120 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999
 C:Accession: T04592
 R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse, T.
 Submitted to the Protein Sequence Database, March 1998
 A:Reference number: Z15378
 A:Accession: T04592
 A:Molecule type: DNA
 A:Residues: 1-221 <BEV>
 A:Cross-references: EMBL:AL022141
 A:Experimental source: cultivar Columbia; BAC clone F23E13
 C:Genetics:
 A:Map position: 4
 A:Note: F23E13.120

Query Match
 Best Local Similarity 78.7%; Score 59; DB 2; Length 221;
 Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGGGSGGRASGGG 13
 DB 98 GGGGSGGRASGGG 110

RESULT 13
 S31415
 glycine-rich protein GRP22 - rape
 C:Species: Brassica napus (rape)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: S31415
 R:Bergeron, D.; Boivin, R.; Baszczyński, C.L.; Bellemare, G.
 Submitted to the EMBL Data Library, August 1992
 A:Description: Characterization and expression of a gene family encoding glycine-rich
 A:Reference number: S31415
 A:Accession: S31415
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-291 <BER>
 A:Cross-references: EMBL:Z15045; NID:g17820; PIDN:CAA78762.1; PID:g17821
 C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match
 Best Local Similarity 78.7%; Score 59; DB 1; Length 291;
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGGGSGGRASGGG 14
 DB 115 GGGGSGGRASGGG 128

RESULT 14
 T20497
 hypothetical protein F02D10.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T20497
 R:Swingburne, J.
 Submitted to the EMBL Data Library, November 1995
 A:Reference number: Z19283
 A:Accession: T20497
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-316 <WIL>
 A:Cross-references: EMBL:Z67990; PIDN:CAA91932.1; GSPDB:GN00028; CESP:F02D10.1
 A:Experimental source: clone F02D10
 C:Genetics:
 A:Gene: CESP:F02D10.1
 A:Map position: X
 A:Introns: 56/3
 C:Superfamily: unassigned collagens

Query Match
 Best Local Similarity 78.7%; Score 59; DB 2; Length 316;
 Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGGGSGGRASGGG 13
 DB 98 GGGGSGGRASGGG 110

RESULT 15
 A33616
 heterogeneous ribonuclear particle protein L - human
 C:Species: Homo sapiens (man)
 C:Date: 30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change 15-Sep-2000

C:Accession: A33616
R:PIN1-Roma, S.; Swanson, M.S.; Gall, J.G.; Dreyfuss, G.
J. Cell Biol. 109, 2575-2587, 1989
A:Title: A novel heterogeneous nuclear RNP protein with a unique distribution on nascent
A:Reference number: A33616; MUID:90078296
A:Accession: A33616
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-558 <PIN>
A:Cross-references: GB:X16135; NID:g32355; PIDN:CAA34261.1; PID:g32356
C:Superfamily: Caenorhabditis elegans hypothetical protein C44B7.2

Query Match 78.7%; Score 59; DB 2; Length 558;
Best Local Similarity 78.6%; Pred. No. 5.4;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GGSGSGGRASGGGS 14
|||||
Db 8 GGSGSGGRYRGGS 21

Search completed: March 15, 2001, 10:52:56
Job time: 1005 sec